DOCUMENT RESUME

ED 347 188 TM 018 662

AUTHOR	Wideen, Marvin F.; And Others
TITLE	External Examinations and the Curriculum: Do They Monitor or Control?
SPONS AGENCY	Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).
PUB DATE	Apr 92
NOTE	52p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Achievement Tests; Curriculum Evaluation; Educational Innovation; Foreign Countries; Graduation Requirements; High Schools; Interviews; Science Education; Secondary School Students; *Secondary School Teachers; Standardized Tests; Student Attitudes; *Teacher Attitudes; Testing Problems; Testing Programs; Test Results; *Test Use
IDENTIFIERS	British Columbia; Curriculum Alignment; External Evaluation; Final Examinations; Large Scale Programs; *Provincial Examinations; Teaching to the Test; *Testing Effects

ABSTRACT

The impact of large-scale testing on curriculum policies at district and school levels and on science teaching at the classroom level was studied for British Columbia (Canada). Other factors affecting teaching practices were considered. In 1983, the Province reintroduced centrally set and graded province-wide examinations for grade 12, following a 10-year period without such examinations. The examinations studied were the grade-12 final examinations in secondary science and science assessments conducted in grade 10. Data were collected from teachers and students in grades 8, 10, and 12, and from principals and district and province administrators. Phase 1 data were collected from 24 teachers in 2 districts. Phase 2 data were collected from observations of and interviews with 56 teachers in 10 districts as well as several administrators. At grades 8 and 10, teachers reported little influence of the government examination. At grade 12, the examinations had enormous impact. Objections usually centered around the narrowing of curriculum, the psychological pressures of the examinations, the erosion of creative teaching, and the notion that some aspects of education were being ignored in science teaching. The grade-12 examinations used in British Columbia are not mere indicators of performance. Evidence indicates that these examinations have effectively frozen innovative practices in grade-12 science teaching. Alternatives to the current situation are discussed. Four figures, 2 tables, and 29 references are included. (SLD)

ED347188

External Examinations and the Curriculum: Do They Monitor or Control?

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
THOMAS O'SHEA
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Marvin F. Wideen
Faculty of Education
Simon Fraser University
Burnaby, BC
V5A 1S6

Thomas O'Shea
Faculty of Education
Simon Fraser University
Burnaby, BC
V5A 1S6

George Ivany
University of Saskatchewan
Saskatoon, SK

Paper presented at the Annual Meeting of the
American Educational Research Association
San Francisco

April 1992

This study was funded by a grant from the
Social Sciences and Humanities Research Council of Canada

BEST COPY AVAILABLE

TM018662

Background

The advantages and disadvantages of external examinations have been argued for decades. Educators, parents, and politicians generally accept the need for some kind of assessment of student progress, but just who should undertake it remains contentious. Behind the seemingly simple notion of determining student achievement lie issues of control over the intended and implemented curriculum and who is to wield power in the educational system.

For those not immersed in the educational system of the United States, the picture is even more clouded. In the Canadian context, it is difficult to understand the furore surrounding such matters as the fall and rise in Scholastic Achievement Tests results or the debate surrounding minimum competency testing. Although our educational infrastructure differs radically from that in the U.S.A., our source of information is primarily American. Furthermore, while the education programs of provinces in Canada are more similar than different, there are important differences in procedures to evaluate individual achievement and to monitor program quality.

In Canada, each province determines its own curriculum and ensures compliance through a central Department or Ministry of Education. Half the provinces (British Columbia, Alberta, Saskatchewan, Quebec, and Newfoundland) require students to pass centrally-set school-leaving examinations as a condition for school graduation. Many provinces also have in place assessment programs to monitor general levels of achievement in various subject areas. Some of these involve standardized tests; others use locally developed curriculum-based instruments.

In this paper, we have differentiated three forms of achievement testing common in Canada. "Syllabus-based tests" are those created to match a clearly defined curriculum and are generally used to assess individual performance for the purpose of school graduation. "Standardized achievement tests" consist of "...a fixed set of test items designed to measure a clearly defined achievement domain, specific directions for administering and scoring the test, and norms based on representative groups of individuals like those for whom the test was designed (Gronlund, 1981, p. 304)." Moreover, on such tests, the items will have been pretested and selected on the basis of difficulty, discriminating power, and relationships to a defined set of specifications, and will in all probability be cast in a multiple-choice format. "Program assessment instruments" are designed to determine general levels of achievement of a cohort of students. Such instruments will be constructed to match a broad curriculum that may extend beyond that covered in school, and may make use of samples of students.

We see three underlying purposes for "large-scale testing," that is, those testing activities that are not under the control of the classroom teacher or the school. Those are selection, accountability, and control. Selection is designed to sort individuals for future education or vocation. Accountability is designed to identify agencies within the educational system that may be held responsible for the failure or success of students, usually as groups rather than individuals. Control centres on the issue of who is to determine what students are to learn and how that will be accomplished.

The Case of British Columbia

If we accept that testing is an inevitable concomitant of instruction in North American schools, what would constitute the ideal situation? Without considering the quality of the tests themselves, each form of test would be used differently. Shepard (1989) cautions that "the differences between accountability and instructional assessment are so fundamental and necessary that it may not be desirable to merge the two purposes" (pp. 7-8). Thus, standardized tests would be used to diagnose individual students' performance, as they were originally intended to do. Such tests would be used, for example, to help make decisions about assigning students to learning assistance or gifted programs. Syllabus-based tests would be used to determine the success of individual students in the school program. Program assessment instruments would be used to evaluate the effectiveness of the educational system as a whole and, perhaps, that of major sub-units.

Shepard (1989) also postulates other desirable characteristics of an effective evaluation program. She points out the need for teachers to have a say in decisions regarding individuals: "Although single teacher tests are probably less reliable (in a statistical sense) than a one-hour standardized test, the accumulation of data gathered about individual pupils in the course of a school year has much more accuracy" (p. 7). For large-scale assessment she recommends the use of matrix sampling, as in the California Assessment, and a "cycle of topics so different subjects could be comprehensively assessed each year" (p. 9).

If one examines the structure of assessment in British Columbia, one might conclude that this province has in place all the recommended ingredients for effective assessment. Let us summarize each of the three testing modes. Although it has been difficult to find specific information on the use of standardized tests in the province, we believe that school districts, in general, use such tests for diagnostic and within-school placement purposes. At worst, districts have a policy to administer standardized tests each year at certain grade levels, but the results are simply entered into the students' files, with no follow-up activity. The lack of information, generally,

would indicate that the results of standardized tests are of little concern to teachers and administrators either for promotion or program evaluation.

Syllabus-based tests are in place for virtually all subjects at the Grade 12 level, and only at that level. Ostensibly, then, the examinations serve the twin purposes of ensuring comparability across schools and reassuring the public that educational standards are being upheld. The Ministry tests contain a mix of multiple-choice and open-ended questions, and the Ministry clearly, and publicly, indicates the examinable body of knowledge through its curriculum guides and tables of specifications for examinations. Moreover, teachers have direct input into the grades awarded because the final mark for the student is based 40 percent on the Ministry examination and 60 percent on the teacher-awarded mark. At all other grade levels, grades and promotion decisions are based entirely on teacher-awarded marks.

The Ministry's assessment program assesses subject areas on a cyclical basis. In Science, assessments occurred in 1978, 1982, 1986, and 1991. The 1991 science assessment was designed to address four aspects of student outcomes: affective, processes and skills, knowledge, and higher level thinking. Although the first three assessments were based on the population of students at the three grade levels using only on multiple-choice items, the 1991 assessment was different. The Ministry now is moving toward alternative means of assessment, including observations and interviews (British Columbia Ministry of Education, 1989). No longer are results to be reported at the school and district levels, and the 1991 Science assessment adopted matrix sampling of items and students to allow greater curriculum coverage at the expense of local reporting.

Structurally, then, the system in British Columbia complies with many recommendations of test critics. But, what is the underlying reality? As previously mentioned, for example, we know little of how standardized test results are used. Are they administered simply because it is a policy to do so? If they become part of a student's permanent record, do they play any direct or indirect role in teachers' or counsellors' recommendations for promotion or program? And why do districts adopt different policies regarding standardized testing?

With regard to final examinations, the conventional wisdom seems to be that teachers teach to the test, at the expense of other instructional goals. Why should this be the case when more than half the student's mark is based on teacher evaluations? In their own evaluations, do teachers use instruments similar to those developed by the Ministry, or do they use greater variety? Do teachers evaluate students on objectives other than knowledge or skills, as Popham (1987) suggests they will using his measurement-driven instruction model? Do they value and teach the processes of developing an understanding of our world through science and attempt to evaluate such

student understanding? Would they do things differently if external examinations did not exist?

With regard to the assessment program in science, we have some information to suggest that teachers have little understanding of how the program operates and do not read the assessment reports. How seriously, then, do students and teachers view the annual assessment program? Are the results reliable and valid indicators of student achievement, particularly at affective and higher cognitive levels? What policy decisions have been made at the provincial, district, and school levels on the basis of assessment results? Have individual teachers been evaluated using these results? Do teachers even care about such matters?

If we are to understand fully the relationship between large-scale testing and science education, it will be necessary to go beyond the rhetoric of advocates and opponents of testing. We have found that much of the critical literature is based simply on the analysis of multiple-choice test items (e.g., Hoffman, 1962; Morgenstern & Renner, 1984). On the other hand, in a circular argument, advocates use test results themselves to justify test-driven programs (see, e.g., Popham et al., 1985). Where the influence of testing on teacher practices is described, that information is based on surveys and interviews (e.g., Darling-Hammond & Wise, 1985).

All these sources are at least one step removed from what actually occurs in the classroom. What is really required is a much closer look at teachers as they conduct their classes, to determine not only their outlook on large-scale testing, but also to understand the role that their own evaluation practices play in their teaching. The study described in this report was developed to explore exactly these issue.

Purpose

The study was designed to address the following questions:

1. What has been the impact of large-scale testing on curriculum policies at the district and school levels and on science teaching practices at the classroom level?
2. What factors in addition to large-scale testing influence teaching practices and the perceptions of the various people involved in the educational system with regard to science teaching?
3. What has been the general impact of large-scale testing in relation to other factors?
4. How can the impact of large-scale testing be explained in terms of social change theory, curriculum decision-making, and the philosophy of science teaching?

Methodology

Province-wide school-leaving examinations had been an integral part of education in British Columbia until 1973 when they were abandoned for a period of 10 years. In 1983, the government reinstated a system of centrally set and marked examinations

> to ensure that grade 12 students meet consistent provincial standards of achievement in the academic subjects. The examination program will also ensure that graduating students from all schools in the province will be treated equitably when applying for admission to universities and other post-secondary institutes. An additional purpose of this program is to respond to strong public concerns for improved standards of education. (British Columbia Ministry of Education, 1983, p. 6)

Given that the province had experienced a ten year period with no final province-wide examinations at the 12th grade level, followed by their reintroduction in 1983, and given that it also had developed a large-scale assessment program, a unique opportunity existed for an impact study of the effects of large-scale testing in British Columbia.

We designed the study initially to examine the impact of large-scale testing on science teaching at the district, school, and classroom levels; to explore the more general effects of a clearly mandated policy initiative on change and decision-making at different levels of a social system; and to address a number of philosophical issues related to the role of large-scale testing in education. Although outside the immediate scope of our initial undertaking, we soon became aware, as a result of work done during the pilot phase, that other important factors, such as teaching styles, district policies, quality of teacher training, availability of inservice, crowded curricula, and teacher background and perceptions of science also affected the quality of science teaching. Moreover, it was clear that it would be very difficult to extricate the effects of large-scale testing from these influences. Hence, to make the study as inclusive as possible, we decided to extend it to include other factors and thereby paint a more complete picture of science teaching at the secondary level. Thus, while the study was specifically focused on the impact of large-scale testing, its scope was broadened to address other research questions and issues.

The central issue of the study concerned the effects of province-wide Grade 12 final examinations in secondary science (Biology, Chemistry, and Physics), and of science assessments conducted in Grade 10. To examine the effects of external testing across the secondary school, we collected data from teachers and students in Grades 8, 10, and 12 and from principals, district staff, and Ministry officials as well. In the first year we conducted intensive case studies of two school districts and in the second year we expanded the study to ten other districts.

Two school districts participated in the first phase of the study. The districts were chosen to represent a large metropolitan school area and a smaller, semi-rural one, and both were amenable to having researchers enter their schools. In this phase, members of the research team observed and audio-taped 4 or 5 science classes of each of 24 teachers (12 in each district; 3 in each of Grades 8, 10, and 12) in 6 schools in one district and 5 in the other. The schools were chosen by district officials, and the principal in each school identified teachers who agreed to allow the research team into their classrooms. We also interviewed various members of the educational community, including the teachers in the classrooms observed, principals and science coordinators in the school, district personnel, and students. Questions on the teacher interviews included the teacher's understanding of science (world view) and science teaching; background, including the number of years he or she had been teaching and his or her preparation in science; teaching style; policy, procedure and practice at the school, and district levels; and the influence of other external factors such as curriculum guides, parental expectations, textbooks, post-secondary institutions' expectations, and the influence of colleagues.

In the second phase of the study, we modified the observation instruments, adopting Stodolsky's (1988) definitions of classroom practice. The interview was refined to capture in more detail the issues that appeared to inform and influence science teaching. To determine the Phase Two sample, the 75 provincial districts were stratified into six geographical regions. Six districts were randomly selected from each region, and four more were drawn from the regions with the largest populations. We contacted the superintendent in each of the ten districts and received permission to approach the various schools in the district. Schools were chosen at random in the larger districts, while in the smaller districts we had to accommodate to the exigencies of school size and the availability of teachers willing to take part in the study. Individual members of the research team observed a single science class of each teacher and followed this observation with an indepth interview with the teacher. Data were obtained from a total of 56 teachers in Phase Two.

In Phase One, each researcher was assigned primary responsibility for a number of teachers. That person undertook the majority of observations for a given teacher, and was responsible for writing up a summative "teacher profile." The profile consisted of a two-part description of the teacher and his or her classes. Based on a printed transcription of the interview audiotape, the researcher constructed a portrait of the teacher under the following headings: background, world view, teaching, influences on teaching, and testing. The second section contained a general description of the classroom followed by a one-page summary of each class observed, including a brief interpretive comment.

In some cases, researchers had spent considerable time in a school and had interacted with department heads, vice-principals, principals, librarians, and other staff members. Where possible, school profiles were constructed along lines similar to, but not as extensive as, the teacher profiles.

In Phase Two, a single researcher interviewed a teacher and observed one of his or her classes. Although the observation and interview were audio-taped, neither was transcribed, and the researcher produced a two-page teacher summary based on his or her analysis of the transactions on the audio-tapes. The observation coding sheets and the interview summaries formed the basis for data analysis in Phase Two.

The wealth of qualitative data contained in the teacher interviews served as the major source to gain insight into the phenomenon of science teaching at an individual level. To paint a more comprehensive picture and search for explanations we developed a procedure to quantify information related to a number of variables prominent in the data. Twelve teacher variables based on the interview questions were chosen. These variables portrayed information about an individual teacher's background and experience as well as that teacher's world view and approaches to teaching:

- number of years of science teaching experience,
- level of science education,
- inservice background,
- world view,
- importance of teaching science in schools
- desired student outcomes,
- approach to teaching science,
- desire for change in science education,
- the influence of external factors on teaching,
- importance of large-scale testing,
- impact of large-scale testing on teaching practices,
- importance of class testing,

For each variable a five-point scale was developed and teachers were placed on the scale based on their responses to interview questions related to these variables. For a number of variables, typical responses were identified during the rating process for each point on the scale. For example, the variable "impact of large scale testing on teaching practices" consisted of five points ranging from "no impact" to "strong impact" with the following typical response associated with each point:

1. "No impact."
2. "The Grade 12 exams have put an additional responsibility on teachers to cover the Grade 8 curriculum."

3. "I give more multiple-choice and fewer rigourous essay questions, and also cover more of the curriculum."
4. "Teaching was more relaxed and there was more time to 'play with' labs before large-scale testing was introduced. Large-scale testing discourages students who have an interest in science but don't want to face government exams."
5. "I used to do an integrated project with the English teacher which is no longer possible. I did more project work, more library work, more open-ended work, and more programs involving the community."

For each teacher we constructed a graphical display in the form of a "star chart" of his or her ratings on all twelve variables by following the general principles laid out in Chambers, Cleveland, Kleiner, and Tukey (1983, pp. 160-162). We gave copies of the star charts for all the teachers to four different members of the research team to sort into groups as they saw fit. The charts did not contain any information as to which variables were represented in each sector nor which teachers were associated with the charts. At the same time we subjected the data to a classical cluster analysis using SPSS and the average linking between groups clustering method. The four raters then met as a team to compare their clusters with that obtained in the computer analysis. The variable labels were revealed and the team, through consensus, identified several subgroups of teachers having well defined characteristics.

Results

This section provides a descriptive summary of the data. We begin with an overview of where the policy change appeared to have its greatest effect. We then provide the data collected relative to science teaching as we observed it in classrooms. We also report on those factors which influence teaching as reported by the teachers in the interviews. Finally, we report the data we collected that directly relates to the question: what has been the impact of large-scale testing on the instructional activity of teachers?

Policy Implications of Large-Scale Testing

It had been assumed that large-scale testing would potentially have effects at the district, school and classroom levels. During the course of our data gathering, and particularly during the first phase in which we worked exclusively in two districts, we sought information about how the introduction of large-scale testing affected policy at those three levels. We looked for evidence that districts or schools undertook some policy initiatives either verbally or in writing that would smooth the transition to final examinations. It soon became evident that we were attempting to document a non-event. Neither districts nor schools produced any evidence of written or spoken policy that had resulted from the introduction of final

examinations. Final examinations it appeared had become a direct link between the Ministry of Education and teachers, with the district and schools acting as observers of the events rather than as participants.

The main role that we saw schools and districts play in connection with final examinations at the Grade 12 level evolved over time and generally took the form of comparative reporting of school and district results. This practice varied from district to district and from school to school. In one school we were shown graphs where each teacher's results on the final examination were represented on a year-to-year basis and the vice-principal routinely shared these data with the teachers involved. In other cases both schools and districts assumed a more laissez-faire attitude toward the results from the examinations, leaving interpretations to the teachers themselves.

While in the field researchers heard about schools which discouraged marginal students from taking Grade 12 science classes in an effort to keep school and district averages high. In keeping with this philosophy we learned of one district in the lower mainland where an enterprising mathematics teacher had encouraged lower ability students to take higher level mathematics courses. While many of these students successfully wrote, others were not as successful and the "erring" teacher was removed from teaching Grade 12 math. As well, in some schools and districts we saw the practice of replicating the final examination scenario at the Grades 10 and 11 levels to parallel those at Grade 12. The rationale here was to "get these students ready for Grade 12."

The most immediate impact, however, occurred at the classroom level. Thus, much of this report deals with the impact of large-scale testing on the work of teachers. Other effects on curriculum and practices, at the schools and districts where they did occur, appeared to be of a more subtle nature as can be seen in the case of the mathematics teacher cited above.

Classroom Practices in Science Teaching

Because the study primarily sought to examine the impact of large-scale testing on instruction, we saw it as important to begin this section with a description of our observations of the teaching we observed and how the teachers saw their own teaching. This section then begins with an aggregated report of the teaching we recorded on timed coding sheets as we observed teachers in their classrooms. The second section reports how the teachers described their own teaching; these data represent a quantifying of qualitative data.

Classroom observations

We first described the science teachers' instructional format in terms of the percentage of time they spent on each of the activities as shown in Table 1. There it can be seen that the most consistent and prevalent mode of instruction consisted of having the students engage in seatwork. Lecturing and instructing were also prominent, particularly in the upper grades. The amount of discussion and laboratory work appeared to decrease

Table 1
Science Teachers' Instructional Format
(mean percentage of class time)

Activity	Grade 8 (n=17)	Grade 10 n=20)	Grade 12 (n=16)
Lecture	3	11	28
Teacher instructing	12	22	29
Discussion	10	4	4
Checking Work	5	12	3
Giving Instruction	14	5	2
Demonstration	2	3	0
Lab	24	8	7
Seatwork	22	27	21
Test or Quiz	2	6	4
Audiovisual	7	3	2
Group Work	0	0	0

with increasing grade levels. However, relying solely on the mean percent of time devoted to each activity across the classes tended to be misleading because the number of classes varied in which the activity was actually observed.

To show more clearly the activities we observed, we cast the information in the form of box-and-whisker plots, as indicated in Figure 1. To interpret the diagrams, one needs to know that the + indicates the location of the median class, the left side of the box indicates the 25th percentile, and the right side of the box the 75th percentile. The dashed line shows the expected range of the upper 25 percent of observations, based on the length of the box. The dot on the dashed line shows the highest observed value within that range. Any values outside that range, that is, to the right of the dashed line are outliers, indicated by x. An extreme outlier is shown as **X**. Where no boxes are shown, fewer than 75% of the classes engaged in that activity, and in such cases the dots indicate the percent of class time for each class.

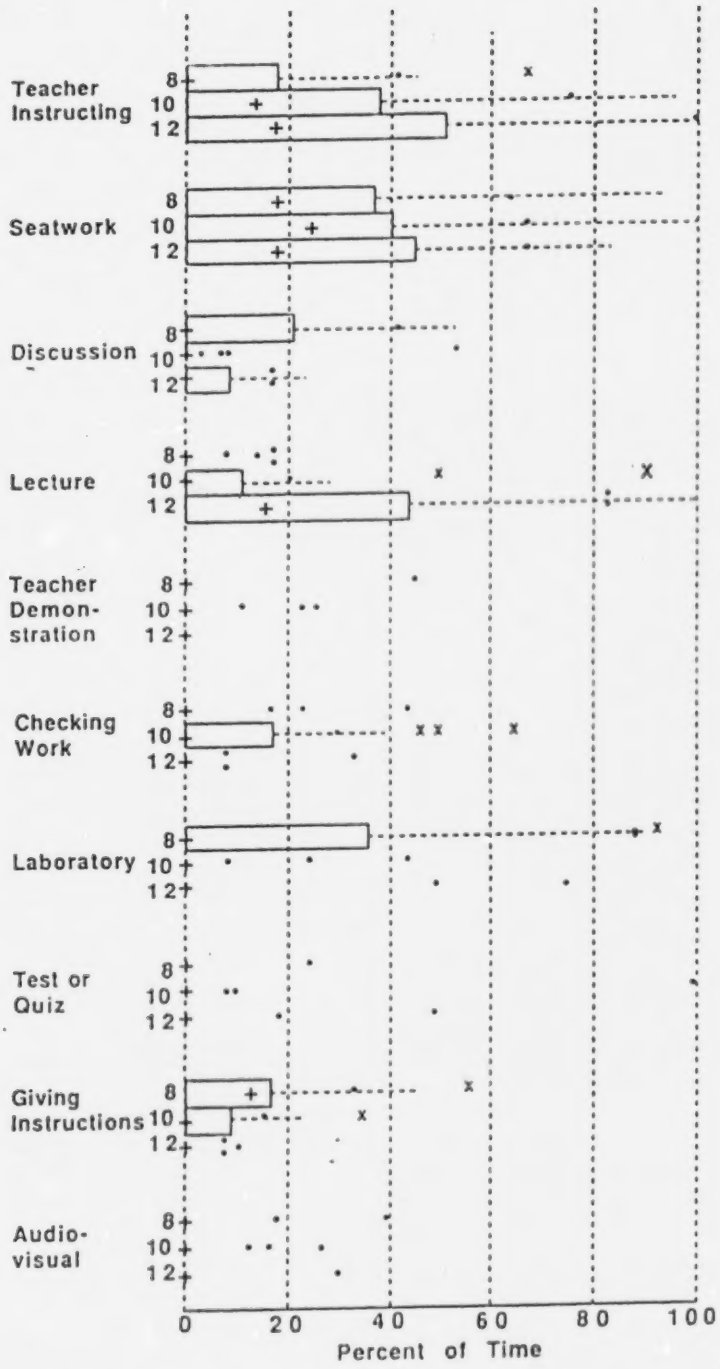

Figure 1: Box-and-whiskers plots showing science teachers' instructional format.

To illustrate, consider the activity "giving instruction" at the Grade 8 level where we observed 17 classes. The median percentage of time teachers engaged in this activity was 13. The first quartile (25th percentile) was zero, and the third was 17 percent. The largest percent within the expected range was 35. In one unusual class, the teacher devoted 55% of the time to giving instructions. The raw data on which the diagram is based consisted of the following percentages: 0, 0, 0, 0, 0, 8, 11, 11, 13, 13, 14, 16, 17, 22, 25, 35, 55.

In general, because no box shows a whisker attached to its left side, at least a quarter of teachers at each grade level did not use each activity. In most cases, because the median coincides with the y-axis, at least 50 percent of teacher did not use that activity.

It is clear, as suggested by the mean values shown in Table 1, that "seatwork," used by over 50% of teachers at each grade level, was the most popular instructional format. In this format, the students worked individually at their desks on a common task. The median percent of time at each grade level was around 20% and the highest values around 70%. In a quarter of the classes we observed, teachers used seatwork for over 40% of the class period.

The next most frequently used format was "teacher instructing" which involved relatively short exchanges between the teacher and students, usually where the teacher called on individual students to answer questions. The use of this mode increased from Grade 8, where more than half the teachers did not use this approach, to Grade 12 with a median of 18%. In one unusual Grade 8 class the teacher used this mode 70% of the time; in Grade 12 one teacher used it for the entire class.

"Lecturing," where the teacher talked most of the time to impart information, ideas or skills, was used by over half the Grade 12 teachers, and for up to 80% of the time in two cases. One extreme Grade 10 teacher lectured for about 90% of the period. Few Grade 8 teachers lectured, and then it was up to about 20% of the time.

A popular mode in instruction at the Grade 8 level, although used by less that 50% of teachers, consisted of "laboratory work." In three cases, virtually the entire class was given to this activity. Few Grade 10 and 12 classes were so engaged. "Giving instructions," where the teacher told the students the plan for an activity, was used by more than half the Grade 8 teachers, and this may be as a result of the higher laboratory activity at this grade level. The incidence of the teacher giving instructions decreased with increasing grade level in seeming inverse proportion to the laboratory work.

The use of "discussion," where the teacher led a discussion and tried to elicit opinions and ideas rather than right answers from students, was more prevalent at Grades 8 and 12 than at Grade 10. At all levels, however, fewer than half the teachers engaged in this mode.

"Checking work," where the teacher provided short answers to homework, seatwork, or tests was more frequently used in Grade 10 than the other grades. Three teachers spent an inordinate amount of class time so engaged.

The three activities of "demonstration," "test or quiz," and "audiovisual" were observed in fewer than a quarter of the classes, perhaps because teachers knew we were coming at a particular time and wished to engage in some kind of "teaching" activity. In no Grade 8, 10, or 12 class did we observe an instance of "group work/tutoring", outside of laboratory work, or of "contests/learning games."

In general, as might be expected, the student activity reflected the instructional format of the teacher. When teachers were lecturing, instructing or demonstrating, students were watching and listening. Similarly, if the instructional format was lab work, seat work, or discussion students were for the most part involved in these activities.

The cognitive level of the activity in classrooms was quite consistent across grade levels. About 55%-58% of time was spent receiving and recalling information or facts (level 1) and 40%-47% of time was spent on concepts and skills (level 2). As for higher order thinking skills such as synthesis and evaluation (level 3), researchers observed no time in Grades 8 and 10 and only about 2% of time at Grade 12 spent at this level.

These results point to a pattern of teaching across the grades that primarily involves lecturing, instructing, and seatwork. Despite that general characterization, some notable differences occurred across the three grades. We saw a narrowing of the instructional pattern as we moved from Grade 8 to Grade 12. As observers, it appeared to us that the most vibrant classes took place at the Grade 8 and 10 levels. Grade 12 classes, on the other hand, were marked with a palpable desire to come to grips with the material presented. A sense of fun and enjoyment seemed lacking here. This was work and made to seem so. We sensed a strong need to process a great deal of material very quickly. The single-mindedness of the enterprise was underscored by the impatience demonstrated by these students when the teacher withheld answers or ventured into territory that would not appear on examinations. They had no time for anything extraneous. In contrast, classes in Grades 8 and 10 showed a more leisurely pace with more time to explore and attempt different approaches. Although the amount of innovation and use of

alternate models of teaching did not appear startling, when we did observe such innovation, we typically observed it at the Grade 8 and 10 levels.

Teacher interviews

The interviews with teachers were analyzed and coded with respect to 12 variables of interest. For example, the variable "teaching approach", was generated from the interview question that asked teachers to talk about what was typical of their science teaching. Teachers who described their teaching as all lecturing or teacher instructing were assigned a 1. The score was incremented by one for each additional method that the teacher described as part of his or her repertoire. For example, a teacher who described his or her teaching as including direct instruction, lab work, projects, co-operative learning and peer tutoring would have been assigned a 5.

The responses of teachers and their subsequent ranking on the "teaching approach" scale described above are shown in Figure 2. These data suggest that there are marked differences in the way Grades 8, 10, and 12 teachers describe their teaching approaches. The majority of the Grade 12 teachers were ranked as 2's. For the most part they describe their teaching as lecture or teacher instruction plus labs. Most of the Grade 10 teachers wereranked as 3's indicating that they describe their teaching as including at least one other method besides the traditional lecture/lab approach. Grade 8 teachers were also mostly 3's but there were significantly more teachers at this level who were in the 4 and 5 categories, that is, they describe a wide variety of teaching methods being used in their classrooms.

The variables, the criteria for assigning numerical ratings to the interview data, and illustrative examples drawn from the data base are as follows:

Science education. Based on the level of formal science education.

Level:	None	Some	BSc	MSc	PhD
Rating:	**1**	**2**	**3**	**4**	**5**

Amount of inservice. Based on the number of inservice sessions a teacher attended in the previous three years and the professional organizations he or she belonged to. Each inservice and organization was given a weight of 1 and ratings were assigned up to a maximum of 5.

Teaching experience. Based on the number of years teaching science.

Years teaching:	0	1-5	6-10	11-15	16+
Ranking:	**1**	**2**	**3**	**4**	**5**

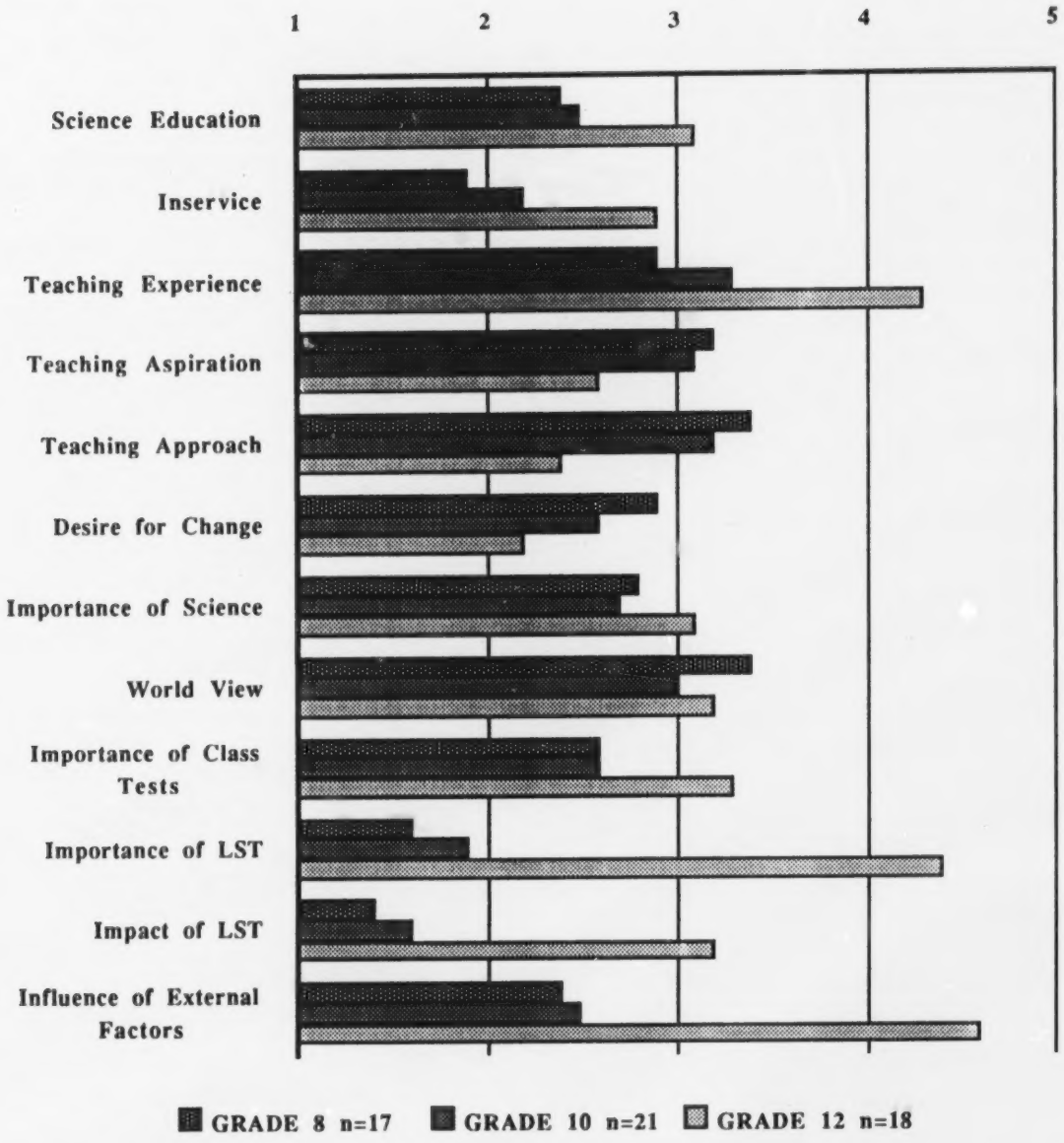

Figure 2: Summary of Phase II Teacher Profiles.

Goals for students. Based on a continuum of "practical" through to "curious". Those who expressed a desire to impart purely practical knowledge were assigned a score of 1 and those who wished that their students leave with a curious nature were given a score of 5.

1. To impart knowledge of the environment and the human body
2. To create an awareness of the technology around them.
3. To create interest and understanding.
4. To develop in students a desire to discover their universe
5. To instil a sense of curiosity about the world.

Teaching approach. Based on a ranking from teacher-centred to student-centred. Teachers were ranked on a continuum from "lecturer" (1) through to "facilitator" (5).

1. Today's class was typical: an hour of lecture with ten minutes of discussion.
2. He teaches with lectures and demonstrations.
3. He divides his class into three equal parts: lecture, demonstration and lab.
4. She uses oral reading, discussion, demonstration, co-operative learning and labs.
5. He does a lot of group work and returns a question with a question rather than just giving students information. He believes that students often have the knowledge and that the teacher's role is to draw it out and help them put it together.

Desire for change. Based on our question: "If you had the power to change anything, what would you do to improve science teaching?"

1. Wants his own video machine and more human anatomy charts.
2. Would like more field trips.
3. The curriculum should be more flexible.
4. Science teaching should be integrated with the other subjects.
5. Get rid of text books, promote hands on investigation, reduce classroom size and integrate the sciences.

The importance of science in the classroom. Based on teachers' responses to our question "what would be lost if science were taken out of the curriculum?".

1. There would be very little lost.
2. Something else would take its place.
3. We could not make rational decisions about how the world ran.
4. Science is necessary for the advancement of society.
5. Devastating to our civilization.

World view. Based on teachers' understanding of the nature of scientific endeavour. Teachers who felt that science is a "search for universal truths" were ranked 1; if they also believed science to be a "set of processes" they were ranked 2. Teachers who felt that science is the "construction of models" were

ranked 5; if they also believed science to be a "body of content" were ranked 4. Those teachers who felt that science was both a "search for universal truths" and the "construction of models" or that science is both a "set of processes" and a "body of content" were ranked 3.

Importance of classroom testing. Based on teachers' views of classroom tests as a means of motivation, review, control, diagnosis, and external accountability. If they scored 3 or more with "very important", we gave them a rank of 5.

1. No use seen as very important.
2. Only one use very important.
3. Two uses very important.
4. Three uses very important.
5. Four or five uses very important.

The importance of large-scale testing. Based on teachers' views of large-scale testing for the same purposes as the previous variable.

1. No factor rated important or very important.
2. One or two important factors; none very important.
3. Three or more important factors; none very important.
4. One or two very important factors.
5. Three or more very important factors.

Influence of external factors. Based on teachers' rankings of six external factors: external examinations, curriculum guides, textbooks, post-secondary institutions' expectations, colleagues, parental expectations. We looked at each teacher's first and second most influential factors. Weights were assigned to each factor as follows: external examinations, 4; curriculum guides, 3; text books, 2; post secondary expectations, 2; colleagues, 1; parents, 1. Ranks were assigned according to the total of the two choices as follows: **1.** three or less; **2.** four; **3.** five; **4.** six; **5.** seven.

The total picture which included all 12 variables is shown in Figure 2. Based on these results, we suggest that a hypothetical teacher who represents the sample would look something as follows. The individual is male, with a bachelor's degree in science (although if teaching Grade 8 he might have a lesser qualification), he has about ten years of experience teaching science, probably double that if he is teaching Grade 12. The total number of inservice sessions he has attended in the previous three years or science teaching organizations he belongs to is around two or three.

He has no well-defined view regarding the nature of the scientific endeavour. While he sees science as a search for universal truths, and those truths make up its content, he also understands science to be a set of processes undertaken to construct models to explain natural phenomena. Consonant with this outlook, he wants to instil in his students both a practical knowledge of science and a drive to satisfy their curiosity in scientific matters, but if teaching Grade 12, he feels the more important aspect is the usefulness

of the knowledge. He sees science as an important subject in the school curriculum, but not absolutely critical for general education; other subjects might be able to take up the slack and generate the skills now taught only in science.

In his approach to teaching science, he claims to strike a balance between being teacher-centred and student-centred. He might say, for example, that he divides his class into three equal parts: demonstration, lecture, and lab. Or that he likes cooperative learning and often uses groups in lab work. If he teaches Grade 12, however, he is more likely to favour lectures and demonstrations, perhaps using lectures rather than discussions to cover the material. He is reasonably satisfied with the status quo. He has adequate resources for teaching, although perhaps would like to have a larger classroom or fewer students or a lab technician to help out in preparing labs and maintaining inventory. If teaching Grade 8, he might like to see broader changes in curriculum such as integrating science with other subjects and increasing the amount of hands-on investigation.

It is only when we consider the issue of testing and external factors that influence teaching and curriculum, that we must make a clear differentiation between Grade 12 teachers and those at the lower grade levels. Grade 12 teachers are much more influenced by external factors of examinations and curriculum guides, and they see large-scale testing as much more important and as having a much greater impact on their teaching. Finally, they view classroom tests as more important than do other teachers. The typical Grade 12 teacher is influenced almost entirely by the curriculum and examinations. Generally speaking, post-secondary expectations, textbooks, colleagues, and parents have little impact. He considers large-scale testing to be very important as a means to motivate students, review material, ensure student attentiveness, diagnose teaching, and ensure external accountability. The impact of large-scale testing has been very strong on his teaching. He may now teach fewer labs and lecture more than before external exams were instituted. He may have previously had students undertake projects in collaboration with other subject-matter teachers. He may feel that he must now concentrate on objectives and reduce the number of side issues students are able to explore or the amount of lab time in which they engage.

Grade 12 teachers were ranked high on all four variables related to testing. Grade 10 teachers as a whole ranked much lower, but still higher than the Grade 8 teachers. The data are consistent in showing the greatest effect at the Grade 12 level and the least at Grade 8. Our sense is that the increase in effect occurs in direct proportion to the imminence of the Grade 12 examination. A confounding factor, of course, is the presence of the Ministry of Education's assessment program at the Grade 10 level. However, assessment in science occurs only once every four or five years, and many teachers could not recall details of the previous occurrence. They seemed to

view it more as a nuisance, taking up class time, and did not think that the students took it very seriously. In looking at the results across the grade levels, we found that 75% of the Grade 8 teachers and 76% of the Grade 10 teachers viewed BCLAP as "not important", while 53% of the Grade 12 teachers felt this way. From teachers' comments it is clear that program assessment is viewed as "low-stakes" rather than the "high stakes" attached to the Grade 12 leaving examinations. And, these findings may also suggest that Grade 12 teachers, although these exams do not directly affect them, are inclined to view testing in general as being more important than do the other two groups.

We also approached the problem of analyzing the data from a perspective other than through grade levels. As described in the methodology section, we undertook a cluster analysis, both by means of the computer and through an inter-ocular inspection of 'star-charts' by four different members of the research team. The star-chart associated with characteristics typical of the sample in Phase II is shown in Figure 3.

When we started looking at individuals within this sample, we soon saw that certain individuals formed sub-groups of their own. Roughly half of our sample (n=27) fell easily into one of the following sub-groups. The group most distinctly different from that shown above consisted of a collection of 11 teachers, of whom 6 taught Grade 8 and 5 taught Grade 10. This we dubbed the "tight" group because the distribution was generally drawn to the centre, as can be seen in Figure 4. In general, they were less experienced, saw science as less important, and their view of science tended to lean toward its practical applications. They were markedly less influenced by external factors in deciding what and how to teach, did not view large-scale testing as important, and felt it had little impact on their teaching. Furthermore, they rated classroom tests as less important than did other teachers. The emergence of this cluster of teachers is congruent with our overall analysis by grade levels. Over a quarter of the Grades 8 and 10 teachers fall into this group.

The second sub-group consisted of five teachers, three in Grade 8 and two in Grade 10, we termed the "bottom heavy " group. In general, they had much less experience than the norm, and their profile showed them to have views on testing similar to those of the tight group. Where they differed was on the variables related to their teaching. Their primary aim in teaching science was to have students develop a sense of curiosity about their lives and their universe. These teachers saw science as a process having as its goal the creation of models to explain the world. Their classes were characterized by an emphasis on student activity and the changes they would like to see were much more extensive than were those of the other teachers and usually concerned major curriculum revisions and improved teaching practice.

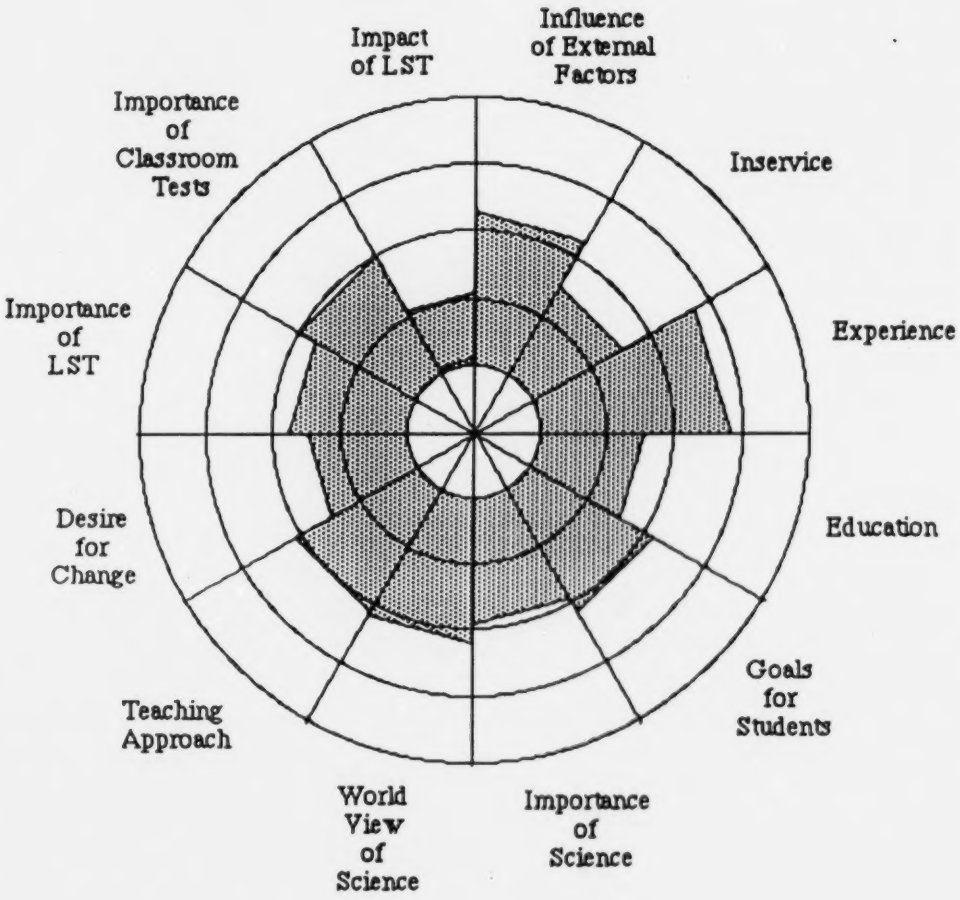

Figure 3. Star chart showing mean scores for 56 phase II teachers on 12 variables

The third cluster, the "top left," comprised a small group of one Grade 10 and three Grade 12 teachers who were more experienced than most. They were more inclined to emphasize practical knowledge and this was consistent with their world view of science as a body of knowledge. They viewed science as an important subject and their classrooms were mainly teacher-centred. They were very strongly influenced by external factors, and the impact of final examinations on their teaching and curriculum selection was very strong. Finally, they considered large-scale testing and their own classroom testing to be very important.

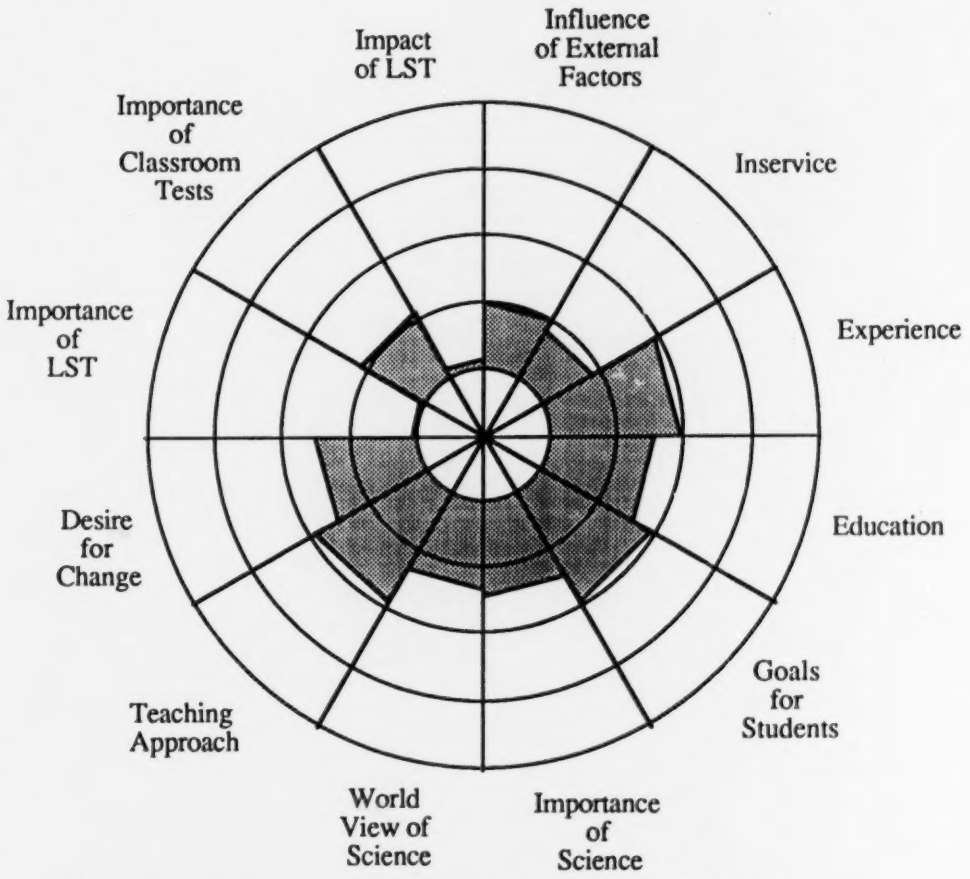

Figure 4. Mean scores for the 11 teachers in the "tight" group.

The final group to emerge, and one that generated considerable discussion among the research team before reaching consensus, consisted of the "group of seven," one at Grade 8, one at Grade 10, and five at Grade 12. In many ways they were similar to the third group. They differed only in two respects; they rated the impact of large-scale testing at the average level, and they felt that classroom testing was less important than average. Nevertheless, they rated external factors as extremely influential in their teaching and they considered final examinations to be an important feature in their teaching.

In synthesizing these results, and looking at the characteristics of the clusters themselves, we noted that the first two clusters were similar in their lack of emphasis on testing. The 16 Grades 8 and 10 teachers in the first two

clusters, in contrast to teachers in the second two clusters, were more inclined to student activity and an emphasis on curiosity. The second two clusters, consisting of 11 teachers, all but three at the Grade 12 level, portray experienced senior teachers who teach a body of knowledge in traditional ways. They are strongly influenced by testing, textbooks, and curriculum guides. The introduction of external examinations may have influenced how they teach or it may have helped to reinforce their preferred teaching style. In either case, they seem to have accepted the system and have little desire to change it.

In general, the cluster analysis and use of the star charts helps to confirm the analysis carried out at each grade level. However, for every teacher who could be grouped, there remains one who could not, and that reminds us that teachers are idiosyncratic beings regardless of grade level. While we can talk about the characteristics of certain groups, we must keep in mind the variability that exists within the teaching community. In the long run, that wide gene pool of knowledge and attitudes may ensure the survival of the species in a sometimes hostile educational environment.

The Impact of Large-Scale Examinations on Teaching

In this section we draw on the interviews we had with teachers, and provide samples of teacher and student comments in secondary classrooms. In analyzing the remarks teachers made about how large-scale testing had influenced their teaching, we found that at the Grades 8 and 10 levels, teachers generally reported little influence of the government exam. However, when we examined the remarks from Grade 12 teachers, we found that the exams had had an enormous impact. Only three teachers in the first phase of our study and two in the second phase reported that they saw merit in the government exam; they made the following statements:

> I think final exams all the way through are a reasonably good idea. People complain that you have to teach to the exam, and you do have to, but I think it's important that there is a reasonable level of consistency throughout the province in terms of what people are taught. (Grade 8 teacher)

> Programs that I'm familiar with prior to that increase in accountability (i.e. before the re-introduction of Grade 12 exams),...had a lot of desirable educational outcomes, but as to whether the programs were achieving it or not was hard to distinguish. So as to accountability, I think that exams are one way of assessing that. Probably the most cost effective way in terms of evaluating the students as well. (Grade 10 teacher)

As can be seen, these teachers value the consistency, accountability and cost effectiveness of the final examinations. While the third teacher who saw some merit in the exams was quite clear in his reasons for doing so, as can be seen from the following quotation, he also felt some misgivings about them, and felt that they had changed his teaching.

> I do know that standardized exams, as much as I hate to admit it, add a certain amount of rigor and a certain amount of determination...(they) provide focus for the students, especially in the later part of the year. However, I don't explore as much as I do in chemistry 11...because I want to direct their attention and these are nice students and I want them to do well in the government exams for their own benefit...a lot of them are really competitive and they do want to do well and you sometimes feel guilty for taking them away from the meat that they need for the course. (Grade 12 teacher)

The two teachers in the second phase who found favour with the government exams both taught Grade 12. One thought that the exams had improved the focus of chemistry 12 by narrowing it to specific topics and levels, and the other felt that both teachers and students were now more accountable.

The response from the majority of teachers, however, ranged from mild ambivalence to strong dislike in their views of the government exams when they were specifically asked, "what impact have Grade 12 exams had on science teaching?" The teachers at the various grade levels who expressed either a mild or a strong dislike for the Grade 12 exams had wide-ranging reasons for their feelings. However, their objections to the exams for the most part centred on four concerns: that the curriculum had narrowed the scope of what was taught; that the various psychological pressures caused by the exams were having an negative effect; that their abilities to teach creatively were being seriously eroded; and finally, that some fundamental aspect of education was being neglected in science teaching.

Concerns about the narrowing of the curriculum touched on several areas. Many of the teachers felt that they no longer taught some topics in as much depth as they once had, others felt that some of the more difficult, but important areas, had been dropped in order to accommodate the exam, and there was a generally agreed perception that much of what was taking place in the classroom amounted to teaching how to write tests rather than teaching science. Among the teachers who held strongly negative views about the exam, there appeared to be a general perception that the curriculum had become too restrictive and had caused teachers to concentrate on core curriculum requirements to the exclusion of anything else. Some of these views can be seen in the following. (All of the comments quoted in this section come from different sources.)

> We have really narrowed our scope, I think. Some people are going to say that we are teaching better because we have to focus on that final exam. I would look at it the other way, that we're probably not teaching as well because we're being forced to do this, this, this, and this and we can never broaden that...I would like the latitude to go further in some areas and that has been taken away. (Grade 12 teacher)

> I think they've had a great deal of impact in that they've permeated all the science courses with a sort of focus to a certain body of information. (Grade 10 teacher)

> I think in the senior levels, the exam is the thing that dictates how you teach and what you teach and how long you spend on it. (Grade 10 teacher)

> (The curriculum) has been narrowed down, dunked down and watered down....They make everything pretty simple now--the physics course is not like it was 10 years ago. (Grade 8 teacher)

Another teacher who refused to teach Grade 12 because of the exams had this to say:

> If I'm back in Grade 11 and 12, I'm really concerned about the examination aspect of it. You cannot branch out, you can't take a topic and expand and go. (Grade 8 teacher)

Along with curricular concerns, certain of the teachers expressed a general unhappiness with what they saw as psychological pressures associated with the final exams. Some teachers felt that the exam produced fear and that this was a poor motivator, others felt that pressures caused by the exams were unfair to both teachers and students, and still others resented the comparative aspects of the exam in which teacher and district seemed to be pitted against one another. This sense of being compared was resented by most teachers who objected to having their teaching assessed largely on the basis of student performance which they felt was not always an accurate reflection of how well teachers had taught a particular course. The following comments illustrate these concerns.

> No I don't like being compared...because we are not the same. The school is not the same, the kids are not the same, the community is not the same....(Grade 12 teacher)

> So you have got that pressure there and it changes how you function as a human being if you've got this thing where everybody is going to be looking at you, how your results are. (Grade 10 & 12 teacher)

> ...in Grade 12 you have to pass those governmental exams and that looms over the students as a negative pressure. (Grade 10 teacher)

Pedagogical considerations also comprised a large proportion of teacher concerns. These teachers felt that changes had taken place all the way down the line; that the Grade 12 exam was having an impact at all grade levels. Some teachers felt that government exams had reduced the opportunities for spontaneity and depth, others complained that because of time constraints imposed by the exams, they were not able to conduct as many labs as they would like. Many of these teachers felt that "the luxury of getting side-tracked" had been taken away from them and that a subtle change had taken place toward a more content oriented delivery which discouraged both student and teacher creativity. Emerging from these comments is a general feeling that too much class time was being spent in preparing students to write the Grade 12 exams and that students themselves "tune out" if they know that what is being discussed will not be tested. One teacher complained

that the exams "reduced the number of unusual and interesting things I am able to do", and several said that they would lecture less if there were no exams. Two of their comments follow:

> The government exams are a B.C. competition game you are playing and you can't afford to be 'down there' just because you went wool-gathering for a month--relevant as that might have been. (Grade 8 and 12 teacher)

> (Before the reintroduction of government exams) I used to do more project work, more library work, more open-ended work and programs involving the community. (Grade 10 teacher)

Another teacher appeared to sum up a general feeling by saying: "...we now learn how to teach towards the exam. In other words, it is teaching us, the teachers."

The above attitudes speak for themselves and must be seen to represent some serious problems in science teaching. But, perhaps the most unsettling concern reported by these teachers involves what they see as an erosion in the kind of science that is being taught. Many of them felt that science classes had become "content oriented" and that they had been reduced to little more than the presentation of a string of facts to be memorized for the final exam. Even teachers who were ambivalent about the exam agreed that since its reintroduction teachers had been forced to "concentrate on objectives, to concentrate on facts." One Grade 12 teacher stated that the government exam had eroded the ideals of good science teaching which included developing a sense of curiosity and a sense of social responsibility. Another felt that in limiting the teacher's ability to cover areas they are keen on, the students miss the enthusiasm and energy that the teacher brings to these topics and which often "rub off" on the students. One teacher summed up his views this way:

> I think they (the government) are trying to upgrade teaching by providing a really severe exam. And, as a result, you have some teachers who maybe aren't that good as teachers but spend the entire year with a bank of exams, going over and over them and the kids maybe haven't learned anything. But, for a certain body of questions they (teachers) are great. But in terms of high level learning and their attitude toward learning and all the other things that education is about, they really aren't good teachers at all. (Grade 10 teacher)

Another interesting aspect to the question of these exams involved the quality of the exams themselves. Teachers appeared to be divided on the subject of whether the largely multiple-choice exam proved to be a fair indication of students' knowledge and whether the exam itself contributed to what many see as a largely content driven curriculum. The teachers who felt that multiple-choice questions were appropriate, did so generally on the basis of multiple-choice exams being easy to mark. Many of these teachers also felt that it was appropriate to teach students how to write exams because they

would be faced with them when they reached university. Those who opposed this type of question felt that the exam often "manipulated students into making process errors"; and that they failed to test the higher cognitive levels. Some of the teachers in this camp also felt that exams motivated students to write exams rather than motivating learning and that they discouraged students who had an interest in science but who did not want to face final exams. As evidence of this last consideration, on several occasions the observers were told about schools or districts that discouraged marginal students from taking science classes in order that the district average might be kept high.

While by far the majority of what we saw taking place in B.C. classrooms, and what was reported to us by the Grade 12 teachers themselves, suggest that final exams are indeed having an enormous impact on science teaching, it would be foolish to ask the reader to assume that we saw nothing else. These occasions were rare, but they did exist. Two of the 16 Grade 12 teachers we observed in the second phase represented exceptions or anomalies to the general perception that the Grade 12 final exams were driving what takes place in the classroom. These two teachers felt so strongly about what they view as good science teaching, that they took it as a point of pride not to let the final exam influence them. In other words, they refused to teach to the test. One of these teachers acknowledged that students in his class, like most Grade 12 students, were unwilling to spend any amount of class time discussing or considering issues that would not be tested. To get over this hurdle, he told the researcher that he makes it a point never to answer the inevitable question, "will this be on the test?" Instead he takes a professional pride in including material that he deems worthwhile and in allowing students time for peripheral discussion around any of the topics raised. He encourages students to ask "dumb" questions. One of his teaching strategies involves opening an astronomy class by telling the students they have 10 minutes to ask him anything they have ever wondered about astronomy. "Well," he said, "I knew what was going to happen, one hour later we were still discussing it." This teacher places considerable importance on teaching a unit on radioactivity, and although it does not appear on the government exam he continues to teach it nonetheless. He told the researcher:

> I have a super lesson on radioactivity which includes some wonderful demos and lab work and great library projects. Because a lot of us don't like to teach it, they don't want it on the government exam. 'I don't care about your exam, I'm going to teach it anyway; my kids will do whatever they do.'

Like the teacher just mentioned, the second teacher also holds strong views on what should be taught in science. He, too, feels a certain pride in doing what he thinks is important. For him this means being willing to forego the certainty of high marks in favour of the luxury of discovery. He continues to teach thermodynamics although it is no longer required by the

curriculum. An excerpt from the researcher's notes after listening to the tape of the interview with this teacher illuminates some of his thinking:

> X claims he had better ability to get students to enjoy science before the advent of exams--he says he hates to see kids stopped from exploring. He feels that the curriculum is the #1 influence on teaching and that it is highly tied to government exams, which he picked as #2 (influence). He says that if he wants to cover certain material he doesn't tell the students that it will be a non-testable item because he knows that if he says, "you will not be tested on this", the kids tune out. He thinks government exams have had an enormous influence on science teaching. He feels they are politically motivated, have taken away some tools that teachers used to use, and give students enormous pressure which he sees as not being conducive to learning. He says he still tries to do what he think important and tries not to be swayed by the exams.

Interestingly both these teachers claim to produce students who achieve excellent results on the government exam. And while they don't feel that the exams have affected their teaching, as can be seen from their comments, they do feel that they have had an impact on science teaching in general and on student attitudes.

Discussion

In attempting to answer the question, "what has been the impact of large-scale testing on the instructional practice of teachers?", we also identified factors other than large-scale testing that had had some influence on teaching. The data showed that final examinations at the Grade 12 level did have a strong effect on teaching at that level. This conclusion comes both from what teachers and students told us and what we observed in classrooms. What meaning then does one ascribe to this impact? In this section of the report, we attempt to address this question.

From our analysis of this question it became evident that the impact of the final examinations currently used in Grade 12 in this province goes well beyond what may have been originally intended by those in the Ministry who initiated them. Whether these effects of the policy of reinstituting final exams represent improvement in the schools or a step backward depends on the perspective one takes about value in education. One thing becomes quite clear from this analysis: the final Grade 12 examinations as they are conducted in the Province of British Columbia are not mere indicators of performance.

We begin the discussion by pointing to what we see as an overall readjustment in the way science teaching, particularly at the Grade 12 level, has been viewed in the province and the effect final examinations have had in the process, using what we call "the pincer movement" as a metaphor to describe what takes place. We then focus more specifically on teaching itself and return to our brief review of science teaching to point to some of the limitations that large-scale testing has placed upon implementing the more innovative practices in science teaching. The third area of discussion moves beyond the immediate classroom to the larger issues of reform in education and places large-scale testing in that arena.

Limits to Curriculum Decision-Making: The Pincer Movement

Different models exist concerning who ought to make curriculum decisions, who determines teaching approaches, and essentially who becomes the final decision-maker in the curriculum process. On the one hand, the objectives of the curriculum, the ways and means of achieving those objectives, and their evaluation could be left entirely to the teacher within the general guidelines set out by the Ministry. Here the teacher can exercise discretion with regards to the objectives of instruction, the ways and means of achieving those objectives, and the evaluation procedures to assess them. This scenario does exist in many jurisdictions in the Western World. In Dutch and Scandinavian countries, for example, teachers do have such

freedom (Van Der Gegt & Knip, 1990). Teachers in many primary schools in our own province perceive themselves to have that freedom, and in fact do exercise it. On the other hand, one could set the learning outcomes and give teachers the responsibility to teach in whatever manner they wish to achieve those objectives.

This latter scenario has driven curriculum makers in the province of British Columbia at the secondary level for the past several years. Learning outcomes have been set by curriculum committees, but the ways of achieving those learning outcomes and their evaluation have been left largely to schools and teachers. The junior secondary programs, for example, provide a curriculum and a textbook which, because of its length, requires the teacher to make choices of topics and to treat them however he or she wishes. The professional autonomy of teachers then remains intact: they have objectives, they have materials from which to choose learning experiences, but how they go about that and how they evaluate their students is left to their own professional judgement. In practice this has led to some diversity, but many teachers appear to follow the textbooks provided quite closely. This scenario, however, has allowed some teachers to try alternate approaches and to experiment with innovative practices as we have seen in this study, in Grade 8 and 10 classrooms in particular.

The policy of instituting final examinations at the Grade 12 level in the Province of British Columbia in 1983 has had an additional effect upon this model of curriculum planning. The fixed model of evaluation imposed at Grade 12 adds another level of constraint on this curriculum model. This move becomes tantamount to creating a pincer movement on curriculum implementation. The fixed learning outcomes provide one part of the pincer and the final examination the other. Teachers now have very little room to manoeuvre either the learning outcomes of the curriculum or the evaluation side. Thus, their movements are restrained on either side leaving little or no freedom for the teacher with regards to the ways and means of instruction.

This added restriction then raises questions about curriculum decision-making and power in education. If one accepts the arguments that have just been made, which our data strongly support, then it becomes clear that teachers in Grade 12 have lost much of their discretion in terms of curriculum decision-making. Final examinations have shifted the power of the curriculum to that group who prepare the final examinations. Teachers are no longer implementors of curriculum, but only those who deliver it.

The further question is whether this development represents a positive move in science teaching or a negative one. On the positive side, it has identified the content on which students leaving high schools should be examined. That was one of the original reasons given for implementing the policy in 1983. On the negative side, the policy has undermined the notion of

teacher as autonomous professional. We also find large-scale testing raising important questions about both teaching and general reform in education, issues to which we turn in the sections that follow.

Science Teaching

We assume that "good" science teaching should reflect different approaches to teaching, authenticity, intrinsic engagement, and a constructivist approach; in general it should aim at developing an educated person. We comment briefly on each of these and discuss how large scale-testing may or may not have furthered those notions.

The curriculum reform movement and developments since have provided educators with an array of teaching models from which learning experiences can be developed. Taken together, they represent more that just a bigger cafeteria of choices. As Ausubel (1963) pointed out some years back, the need exists for different types of learning processes because each promotes different educational objectives, all of which become necessary to the process of education. Problems, he argued, can arise with any approach to instruction, not because it is inherently bad, but because it is used for the wrong purposes. Joyce and Weil (1980) have since developed that argument much further through their work around models of teaching. They argue against dogmatism in style, positing that competence in teaching comes through extending one's ability to use different models of teaching. They base this argument on the notion that no single model can achieve the range of objectives thought to be necessary or expected. Like others, they argue, for example, that student understanding of the heuristics of a discipline and the development of attitudes are not necessarily achieved through a didactic approach.

Large-scale testing does not encourage teachers to use different approaches to teaching. It appears to have narrowed the range of instructional practices being used in Grade 12 compared to Grades 8 and 10, and indeed it may also have narrowed the purposes of education itself. Three things appear to be happening as a result of large-scale testing which work against the use of different approaches to teaching. The first involves a time factor. Teachers regularly reported not having time to use any strategy other than the direct teaching we reported earlier. Field trips, for example, simply took too much time away from class; teachers reported having reduced such trips since final examinations had been brought back. Even something as simple as introducing the use of hand-held calculators was frowned upon because time did not permit . Given this restraint, it becomes difficult to imagine a teacher wanting to take the time to experiment with co-operative learning or inquiry teaching. Second, the students do not wish to undertake any activity that would distract them from the task of preparing for the examinations. Earlier we spoke of the "no nonsense" atmosphere in most

Grade 12 classrooms we saw. Although we may see this attitude as a positive factor because it produces on-task behaviour, we can also see that it produces a focus on the content of the curriculum to be examined at the end of the year. Other outcomes, which may also be an important part of the education of children, take on relatively less importance when teachers are encouraged to focus on the content of the curriculum. A third factor involved the examinations themselves; success on them depends primarily upon committing to memory the content of the curriculum and algorithms required to solve set problems likely to appear on the exam. This technique underscores the mechanics of test writing rather than promoting any clear understanding of science concepts.

One of the most important contributions of the curriculum reform movement of the 1950s and 1960s was the emphasis on the authenticity of the disciplines themselves. The wholesale revision of science came in part from a reaction to the pre 1950s' science curricula which were mostly textbook driven and based on a particularized view of science held by teachers and textbook writers rather than by scientists. The revision which occurred saw scientists themselves taking centre stage in the reform. Two factors typified the curricula that emerged such as BSCS, ESCP and PSSC. First, we saw a move towards the use of the "inquiry method" which became the cornerstone of those curricula. Students were encouraged to ask questions, undertake investigations, and seek answers to questions. This way students would come to know something about how knowledge came about as well as becoming familiar with the knowledge itself. Shulman and Tamir (1973) set out four levels of inquiry that could take place in a laboratory which are depicted in Table 2. The level of inquiry becomes determined by the degree of openness that a teacher allows. At level one, the problem, ways and means, and conclusions are all given. Different types of inquiry results as the teacher leaves successive levels open. The proponents did not intend that the move to more inquiry teaching would replace the need to understand and learn content. Rather, the two should go hand-in-hand in a natural way and reflect the authentic nature of the discipline involved.

Table 2
Levels of Openness in the Teaching of Inquiry (Shulman & Tamir, 1973)

Level	Problem	Ways & Means	Answers
One	Given	Given	Given
Two	Given	Given	Open
Three	Given	Open	Open
Four	Open	Open	Open

The second aspect of the discipline that became central to the curricula packages which came from the curriculum reform movement involved what Bruner (1961) initially termed the "structure" of the discipline and what others have since come to call "understanding". Bruner (1961) made the claim that any discipline could be made comprehensible to any child at any level. Philosophers have more recently talked about the need for knowledge and understanding in the disciplines. The central theme running through these ideas posits that to understand a discipline one must not only know something of the knowledge of that discipline but also how that knowledge came about and how it fits together in ways that allow us to understand the discipline.

How has the reintroduction of final examinations at the Grade 12 level affected the authentic nature of science teaching? We saw little inquiry teaching in the Grade 12 classrooms we observed, and only a few instances in Grades 8 and 10. Most of the teaching we saw occurred at level one in Shulman's depiction of levels of openness. One can hardly attribute this situation to the existence of final Grade 12 examinations alone. Many other factors augur against the use of inquiry teaching in classrooms. We identified several of these during our observations and interviews. The curriculum itself, with its heavy emphasis on content, did not appear to encourage the use of this particular strategy. We also gained a sense from our interviews that as a strategy, inquiry teaching had not been part of the vernacular of teaching over the last few years. At the district level, we found that in the few districts where any consideration was being given to science teaching, the emphasis was on building a pool of items with which to improve assessment. We could speculate that the reintroduction of final examinations has changed the professional agenda of the teaching force regarding how they maintain and improve their teaching. In the 1950s and 1960s that agenda focused upon the curriculum and the teaching of that curriculum. Today we see the use of alternate teaching strategies such as inquiry which had been part of that agenda set aside as testing and evaluation have taken over.

The question as to whether an emphasis on examinations has improved students' understanding in terms of knowing the structure of the discipline in the sense that Bruner (1961) described it, remains open to debate. Our analysis produces mixed views. On the one hand, Grade 12 testing appears to have focused the attention of both teachers and students on a specified area of content, and the degree to which the examinations reward that understanding is the degree to which students are assumed to understand the subject. This argument is countered by the observation that examinations also particularize knowledge to an extent that the isolated parts are rarely brought together. The time to sit back, contemplate, and come to understand what a subject area means does not appear to be occurring in the present Grade 12 classrooms. In the words of one Grade 12 student,

> To understand physics you need to play with it, and have fun with it. In school we never do that.

This student, an award winner, reflected on the loss that seems so essential to the understanding of subject matter which is the playful, contemplative side of learning. We saw very little evidence of this side in Grade 12 classrooms; we saw some evidence of it in Grade 8 and Grade 10 classrooms.

One of the notions that received attention during the period of curriculum reform, and one that has been emphasized since, involves student engagement. The concept takes various forms. First, there was the notion of on-task behaviour that emerged from the research on teaching movement; those students who remain on-task produce better achievement. More recently, the concept of active learning has come to the forefront. At one time in science that concept meant 'hands-on' activity; if students were handling materials that was better than if they were listening. This rather simplistic notion has given rise to a more thoughtful view of of active learning that does include hands-on activity, but in addition encourages engagement of students in classroom discourse and even in the selection of learning activities in some cases. A quotation from the Professional Standards for Teaching Mathematics (1991) quite adequately conveys the current meaning of student engagement in classroom discourse:

> Students should engage in making conjectures, proposing approaches and solutions to problems, and arguing about the validity of particular claims. They should learn to verify, revise and discard claims on the basis of mathematical evidence and use a variety of mathematical tools. Whether working in small or large groups, they should be the audience of another's comments--that is they should speak to one another, aiming to convince or to question their peers. Above all, the discourse should be focused on making sense of mathematical ideas, on using mathematical ideas sensibly in setting up and solving problems.(p 45)

These comments about discourse in mathematics apply equally to science teaching; in fact they could have been drawn from Piaget's work in the 1950s. This statement, however, does not mention the notion of having students take some role in selecting their own topics for study.

The question as to how the introduction of final Grade 12 examinations affected student engagement becomes a difficult one to answer. In some Grade 12 classrooms, because of the press of content to be learned, students were found to be on-task in ways that were not evident in Grades 8 and 10. However, since Grade 12 students are a select group, that outcome may have been expected. What we did not see in most of the classrooms to any great extent in the lecture/discussion part of the lessons were discussions and arguments among students about the concepts they were learning. Verification remained something for the teacher to do; students became the

audience, not of each other, but of the teacher. Virtually no student-student interaction occurred in the classrooms we observed. The laboratories provided a different opportunity and setting; but here too, the dialogue centred around verification of well programmed results. Where we did find exceptions they generally occurred in Grades 8 and 10.

Many contend that constructivism as it applies to science in the 1990s bears many of the characteristics of inquiry learning of the 1960s. We both agree and disagree. Certainly, many of the instructional approaches carry strong similarities; the general intentions appear somewhat similar. But we find the constructivist to offer a much more grounded position and in some ways a more radical one, depending on how far one wishes to carry the concept. While inquiry learning of the 1950s typically provided a more meaningful way of developing the conceptual side of the discipline, constructivism frequently challenges its foundation and how it is taught.

MacKinnon (1989) recently summarized three principles that follow for science teaching from constructivism. First, he contends that teachers must first develop strategies that will allow them to become aware of the ideas held by their students about science concepts as these are derived from their ordinary language. Second, these ideas must be taken into account during instruction so as to construct new concepts. Third, students should be actively involved in the learning situation in ways that we describe in the previous section. Under this view, learning shifts from the 'empty vessel' notion to conceptual change which occurs only when students experience dissatisfaction with existing concepts through anomalies and puzzles, develop new conceptions, and see the plausibility of these. Constructivism points to the difficulty of learning new scientific concepts because one's ordinary explanations of phenomena usually suffice.

We saw little evidence that teachers had any awareness of the notions of constructivism. Moreover, while some remnants of the three principles just outlined did appear in Grades 8 and 10, the teaching in most Grade 12 classrooms appeared based on a very different model of teaching and learning from those suggested by MacKinnon. It appeared that the authority for knowing and accepting information came primarily from the teacher and the textbook. Conceptual change did not enter into the picture; in fact, pausing to become aware of the students' conceptions of phenomena became problematic because it took too much time. In conversations with a teacher who attempted to work from a constructivist approach in Grade 12, he cited the students' negative reaction to it as a major hurdle. Basically, students wanted the factual information that would be useful on the examination; they did not wish to be bothered with understanding.

Change in the Schools

How change occurs within a social system has occupied a number of social scientists over the years. The work of two groups provides a background for the discussion that follows. Baldridge (1972) has contrasted two perspectives: human relations and political systems. The human relations perspective is social-psychological in nature, focusing on the individual and peer group relations. Chinn and Benne's (1976) description of "normative reeducative" approach to change illustrates this perspective. It assumes persons to be inherently active, in pursuit of impulse and needs satisfaction; change here comes through altering norms and acquiring changed understandings. The human relations approach to change reflects what has commonly been referred to as the "bottom up" approach to change in a social system. Teachers seeking to make change within their own school would fall under this general approach. Baldridge contrasts this with the political systems approach to change which emphasizes the dynamic features of the organization, assumes that conflict is natural, that power elites govern most decisions, and that external groups have a great deal of influence. The introduction of large-scale testing examined in this study represented a political perspectives approach to change, one which involved the exercise of power as a means to effect change. It came despite little evidence that it was needed and against the wishes of the teachers' organization in the province. The question of interest in this study was how the districts, schools, and teachers changed to accommodate the introduction of that policy.

This study provided an opportunity to examine both the change process itself as well as some of the actual changes the were brought about by the policy decision to reintroduce Grade 12 final examinations.

The introduction of large-scale testing came as a mandated change in the province. To use the terms of Bennis, Benne, Chin & Corey (1976), it came as a "power coercive" move on the part of the Ministry where teachers had little choice but to comply with the initiative unless they decided to change the grade they taught. Once in place, however, certain other effects began to be experienced. The preparation and marking of the examinations began to be seen with some favour among teachers who described that experience as wonderful inservice for them, and also it gave them the inside track for preparing their students for the coming year. The camaraderie and networking that developed around the technology of processing the examinations had a positive effect upon the teachers involved quite apart from any effects upon their instructional practices. The policy decision also had an effect upon how teachers taught science, and how they began to view professional development.

But as we indicated earlier in this section, the final Grade 12 examinations have become more than mere indicators of student performance; their effect can be felt beyond the classroom, raising issues of a broader educational nature. The issue of social change can be see also if we take a more global perspective. For this examination, we need to step back and look at the high school in a more general and global way.

A research team from Simon Fraser University recently reported the results of an extensive review of literature and a follow-up study of 16 high schools in the District of Surrey (Wideen, Pye, Naylor & Crofton, 1990). The literature review pointed to the high schools as an area in considerable difficulty with high dropout rates and increasing numbers of students feeling at risk. The dropout rate between Grades 8 to 12 ranged from 30% to 60 % among schools in British Columbia. The follow-up study to that literature review pointed to the teaching that takes place in the secondary school as the major problem for those students who dropped out and also for those who remained. Lecturing, seatwork and recitation dominated the textbook driven instruction in those 16 secondary schools. The question that master planners will have to face soon is just how reform can be brought about in the high schools. Within this context we then place the practice of final Grade 12 examinations. Both our data and our interpretations of that data suggest that final examinations are having the effect of effectively freezing current practices at the Grade 12 level. The opportunity or motivation for teachers to attempt changes to their practice, a necessary condition some would posit for reform, is effectively nullified by the practice of final examinations.

We make this claim based on two observations, the first coming from what people told us about the impact of final exams, and the second from what we observed in the way of teaching. As we noted earlier, several teachers at the Grades 8 and 10 level reported that the Grade 12 examination had little or no effect on their teaching. Grade 12 teachers consistently told us the opposite, that such examinations were affecting what they did in the classroom. From the data collected during classroom observations, we would be inclined to agree with these teachers. When we observed their practice, almost without exception, any indication of exemplary science teaching that we witnessed took place in the lower grades. For example, in one Grade 8 class, the making of peanut butter was used to demonstrate chemical change; in a class in genetics, Grade 10 students explored changes in a developing fetus by writing essays in which the students viewed the process through the senses of the fetus. On the other hand, the majority of straight lecturing and non-student involvement in class took place in Grade 12 classrooms. An emphasis on tests and testing appeared to dominate many of these classes and the most prevalent student question we heard there was, "will this be on the exam?" It seemed to be a common practice that the last three or four weeks of

Grade 12 classes would be devoted to going over old exams and attempting to review and predict what the form and substance might be for the coming one.

If our contention that Grade 12 examinations have effectively frozen innovative practice in Grade 12, then it can be argued that that fact will in the long run augur against improved practice in the high school. This contention rests on the belief that such improvement can only come from teachers who feel some freedom to experiment with their practice. Here a normative/reeducative approach to change as described by Chinn and Benne (1976) would ensue. In the present system where the main preoccupation appears to be content coverage and a focus upon the evaluation instruments themselves, it is very difficult to imagine anyone venturing into any type of innovative practice. For some this situation might not suggest a problem. But if you take the view that the high school presently is an institution badly in need of reform (see for example, Goodlad, 1983; Sewall, 1983; Van Til, 1976) and further speculate that teachers require some freedom to experiment and take risks to bring about this reform, then the existence of large-scale testing effectively reduces the chances that such change will occur.

Alternatives to Existing Assessment Practice in British Columbia

This study sought to assess the impact of large-scale testing on science instruction. In doing that, it has pointed to a number of undesirable consequences of high school leaving examinations. They appear to be driving or justifying a type of instruction typical of the pre-curriculum reform period of the 1960s; they appear to be standing in the way of changes in the high schools, changes that one would hope would reduce the incidence of students dropping out. The question arises, how should these findings influence policy-making with regard to assessment and the curriculum? The temptation is strong to take a definite position either for the abolition or retention of school leaving examinations, and discuss that position in the light of these data. But such a stand would not reflect the intent of this study. Rather, we find it more appropriate at this stage to set out the alternatives in such a way as to extend and inform the debate that will occur as we see the forces of relevance and constructivism coming up against the pressure for national testing both in Canada and the United States. But if that debate occurs only at the level of whether large-scale testing is good or bad, it misses the point. Such debate must occur within the context of what constitutes a good education and the complexities that surround our schools. In short, whatever policies we undertake must be supported by argument and stand up to debate; mere political expediency will not do.

To encourage such debate, we offer three alternatives to the existing assessment practice in British Columbia along with a fourth, to retain the status quo. We call the first alternative the curriculum option; it seeks to

shift the emphasis away from large-scale testing and toward the improvement of instruction. The second involves changing the assessment process itself as a means toward improving the curriculum. The third suggests that school-leaving examinations be eliminated entirely. These three alternatives make a common assumption, that the status quo with regards to assessment is not good enough. A fourth option does not make that assumption and suggests we leave things as they are.

The following hypothetical scenarios are based on the preceding options. We ask the reader to let the curtain rise on the different options, consider their merits and defects, and perhaps as an astute reviewer would, form some judgement on which scenario holds out the most promise for improving what we understand when we use the term "a good science education."

Switch humps on the camel: The curriculum option

Scenario A. In this scenario the effects of large-scale testing on the instructional activity of teachers have now become known through various studies such as the one described in this report. Those whose interests lie primarily in curriculum have used such information to refurbish their arguments against the excessive emphasis on testing. Other developments also push these arguments to the forefront. Supporting documents for the recent Royal Commission on Education in British Columbia (Sullivan, 1988) have raised serious questions about final mandatory examinations. Policy makers have yet another concern. Large numbers of students drop out of school, and the rate appears to be increasing. The heavy emphasis on examinations at the Grade 12 level appears to be contributing to the dropout rate (Naylor, 1990). But the dropout rate is only a symptom of a much more serious problem--by most standards the high school as an institution finds itself in much difficulty.

Reports point to the curriculum and the way it is taught as the area which most requires improvement. Recognizing the problem, and experiencing some major concerns about productivity, a newly elected government seizes upon curriculum reform as the area in which to support various initiatives that they see making a difference.

The arguments made by curriculum people and by policy makers that eventually win the day hold that final Grade 12 leaving examinations should be maintained but only if they become one of many tools for assessment rather than something that drives instruction as is currently the case. In this scenario, the policy makers draw on existing agencies within the Ministry to focus upon improvements both in the curriculum itself and the way teachers offer that curriculum to students. The curriculum is brought closer to a more authentic notion of science than that found in the present measurement

driven curriculum. Efforts are also made to shape a curriculum that appeals to an wider audience of students at the secondary level than that currently the case in British Columbia.

Scenario B. This scenario rests on the same assumptions as the first, namely that curriculum improvement and instruction must become the focus for reform in the secondary schools. In this scenario, the Ministry can not provide the innovativeness required to make significant changes. The emphasis shifts to quasi-independent groups or consortia made up of schools, districts, colleges, universities, and other organizations to undertake curriculum development projects in ways similar to that which occurred in the curriculum development reform movement in the 1960s. The Ministry funds such projects using in part the monies saved from a reduced emphasis on testing. The testing branch in the Ministry now enters a period of benign neglect.

Build a better mousetrap: Improve assessment practice

Scenario C: In this scenario school leaving examinations are here to stay. The government believes it has the support of the public for what it does and it wants to continue to give assurances that it is "maintaining standards." Moreover, moves are afoot to propose a national curriculum and a common examination across provinces. Clearly, the political winds favour large-scale testing both as a means to reassure the public that it is getting its money's worth and as a device to ensure that teachers at all levels adhere to the centrally-set curriculum.

On the other hand, senior level educators within the Ministry are aware of reports that have found undesirable consequences of testing within classrooms that they would hope to avoid. They are generally in favour of maintaining some kind of metre-stick to assure the reliability of school-leaving marks so employers and post-secondary institutions may have some convenient cut-off by which to assess prospective entrants. At the same time, they wish that teachers would reconsider their heavy reliance on lectures, textbooks, and large-group teaching that characterizes teaching at the senior grades. They seize upon Popham's notion of measurement-driven instruction.

They agree that in order to change teaching practice, you must change assessment practice. If teachers are to teach to the test, change the test to bring it into line with current thinking about what is valuable in education. They consider what needs to be done to create tests that fall under the rubric "authentic" assessment. At the present time, examination items are partly multiple-choice, partly open-ended. Most items test lower-level thinking, in spite of the official table of specifications. Students must answer all items,

with the possible exceptions of blocks of items relating to optional sections of the curriculum.

The test developers restructure the examination to include a set of multiple-choice items that test basic concepts. However, they recognize that items are only samples from the population of possible items, and they decide to allow students to show what they know, rather than restricting them to a set of items that test developers think they should know. To do this, they set a total of 50 multiple-choice items, and ask students to answer any 30, or direct them to answer any 3 out of each block of 5. As justification for this departure from established practice, they cite the long-standing elective procedures used by English examining authorities such as the Cambridge Examination Syndicate. They also instruct their computer programmers to take into account the implications for machine-scoring programs.

To move teachers away from simple objective testing, they include a question that requires students to write a report on an original investigation they conducted during the school year. They might have been given a number of potential topics during the year, or students might identify their own. To ensure the work has actually been done by the student, the teacher might be asked to corroborate the report, before returning the completed examination.

To challenge the very good students they include a bonus question worth the square root of the difference between the mark obtained on the rest of the examination and the total possible mark. The question is one students have not seen before and integrates knowledge across the usual sub-domains within a subject.

Open-ended items are set that do test higher-order thinkings skills. These are marked holistically rather than analytically to reduce the time required for marking. To allow students adequate time to respond, the time limits for tests are relaxed.

Scenario D: In this scenario the assumptions are the same as for Scenario C. This time, however, the decision is made to be more prescriptive of what teachers will actually do in their classrooms. The centrally-set examination will test only basic concepts and knowledge and will do so entirely using a multiple-choice format. Students may have a choice of items. Teachers are then informed that they may not use multiple-choice or simple response items in their own evaluation. They are now expected to deal with the higher-level objectives and this may be done through student investigations, laboratory work, field trips, cooperative group work, individual projects, and the like. Moreover, they are now required to justify the mark they assign to each student that is weighted 60% in the final grade. They are required to submit a teacher dossier containing work that a sample

of students have done during the year, and a record of how the teacher has evaluated that work. At the same time, it is made clear to educational authorities that there need not necessarily be a positive correlation between marks achieved by the student on Ministry examinations and those awarded by the teacher. If we are to believe that rhetoric respecting different student learning styles and interests, there is no reason to believe that the two sets of marks should be closely related.

The task of the marking committees then becomes one of assessing the adequacy of the assignments and the appropriateness of the way in which the teacher marked the work. In numerical terms, the reduction in marking is by a factor of 25 to 30. The highly vaunted inservice component attested to by teachers on marking committees then becomes real inservice rather than simply a chance to get teachers together to discuss how marks should be awarded to students and to compare students' responses. Now they have an opportunity to see and read what their colleagues are actually doing. They are exposed to innovations undertaken by other teachers, and have a standard by which to evaluate their own teaching practice.

Scenario E: This is a variant of Scenario D. In this instance, the student is obliged to submit a portfolio of assignments along with the final examination. The assignments are intended to engage the students in research projects, laboratory investigations rather than the pre-cooked activities of textbooks, and other forms of open search. The type of assignments may be set in the curriculum or they may be chosen by the teacher. The evaluation scheme to be used is determined in advance in the curriculum guide or it may have been set up by the individual teacher. The markers select a random sample of assignments and evaluate them independently of the teacher's evaluation. They compare their evaluations with those of the teacher. If there is reasonable agreement, they accept the remaining teacher grades as given. If there is consistent deviation, they may adjust the teacher grades to suit. If there is no consistent pattern, they may examine all portfolios and assess individual grades. In each case, a report would be submitted to the individual teacher so he or she may make the necessary adjustments the following year.

Scenario F: The same well-meaning educators are at work under the same set of political constraints. This time, however, they focus on the issue of the appropriateness of the examination for all students. They are aware that one cannot set a single fair examination for students of different abilities. For bright students the examination will not be challenging enough for them to show how well they have mastered the subject. For slow students, the examination experience is likely to provoke anxiety and aversion. Part of the problem has been overcome by setting a second "scholarship" examination for students intending to attend post-secondary institutions. However, even here, the results on the scholarship examination are not used to select

entrants to institutions but rather to award scholarships should students wish to continue their education.

The examiners decide to implement a system based on changes taking place at the present time in England. They set a 2-hour examination consisting of four sections, say, W, X, Y, and Z. The items in each section are qualitatively different; W contains the easiest and Z the most difficult. They also differ in the level to which they test higher-order thinking. Students choose just two contiguous sections and have one hour to write each. The available grades range from a high of A to a low of G. Anything less is simply no credit. They know, however, if they choose sections W and X, the only grades available to them are E, F, and G, but they are also likely aware that they have little chance of achieving anything higher in any case based on the feedback they have received from their teacher during the year. Should they choose sections X and Y they may obtain a C, D, or E (or in exceptional cases, an F) A choice of sections Y and Z would yield a grade of A, B, or C (or in exceptional cases, a D). The exceptional cases may consist of students who appeal an award of no credit and who can substantiate their appeal with evidence of being poorly advised about attempting a given level, or for other reasons.

In this way, students can set their own goals for a course. Teachers can structure educational experiences appropriate to the student's aspirations. The hurdle at the end of the race track is adjusted to suit the stature of the student. The educational goal of continuous individual progress is seriously addressed.

Cut our losses: Eliminate school-leaving examinations

Scenario G: In this scenario, government exams are abolished. Curriculum guides are maintained but certain areas of the curriculum are made optional in order to reduce the amount of the curriculum required to be covered. Teachers are encouraged to take risks in the classroom, to consider innovative teaching methods, and to cater to the interests and different learning styles of their students.

Evaluation of teachers is left to the school and district staff and is based on a model of monitored professional autonomy. During teacher evaluation by an administrator, open discussions occur about the ways in which the teacher being evaluated is covering the core curriculum and some parts of the optional curriculum, as well as the way in which the teacher conducts student evaluation. In this scenario teachers feel a renewed sense of professional autonomy and a sense that what takes place in the classroom represents a shared responsibility of teachers and students.

Teachers evaluate students using an eclectic approach which might include a portfolio of their written work, traditional testing, oral presentations, class discussion, debates, and peer evaluations. This approach would evaluate students within their own school context and provide more equity in evaluation because it would not give unfair advantage to students who are good test writers or to students from schools that concentrate on producing good test writers.

Let sleeping dogs lie: Retain the status quo

Scenario H: The arguments against large-scale testing diminish in force as a national proposal for a nation-wide testing program gains popularity. The arguments of cost effectiveness and maintaining standards prevail and school-leaving examinations continue as before. Grade 12 subjects that do not have a provincial examination component, such as Survey Mathematics 12, are viewed as second-rate because they are not "examinable." The universities decide not to accept such subjects for university entrance. As a result, the Ministry sets examinations for all Grade 12 subjects.

The pressure for performance in the examinable subjects works its way down the system. Districts impose district-wide final examinations in Grade 11 so that students will have a dress rehearsal for their final year. The demand grows that the Ministry provide, or fund the development of, suitable test item banks for Grade 11 subjects. The lure of using computers as testing instruments grows, and the Ministry finances a major thrust toward individualized testing, using large item banks calibrated using latent trait models. The use of multiple-choice items grows as more reliance is placed on the computer. The Social Sciences and Humanities Research Council continues to fund studies that examine the impact of large-scale testing on educational practice.

Postscript

In the time between the completion of our study and the final revisions to this report, three Canadian studies, two from Alberta and one from British Columbia, have come to our attention. We feel that these studies are important, and that they help illuminate our findings and confirm our conclusions.

The first is a draft of a doctoral dissertation (Puhlmann, 1991) that examined the effect of introducing school-leaving examinations, or diploma examinations as they are known in Alberta, on the teaching of trigonometry at the Grade 12 level. Puhlmann studied intensively the classroom practices of three Grade 12 teachers, and surveyed an additional six. He spent 6 to 10 periods in each teacher's classroom, and interviewed both students and teachers. His concluding comments include the following:

> The deviations from the intended curriculum and the examples of questionable classroom practices seem to be the result of a vicious cycle which consists of reducing course content to its most manageable and measurable fragments, splitting the learning process into means and ends, and emphasizing the province-wide exam as the most important event for teachers and students....The result is that students acquiesce to the pattern of classroom knowledge thereby participating in their own de-skilling of mathematical knowledge (p. 261).

The second is a study commissioned by the Alberta Teachers' Association to assess the impact of diploma examinations on the teaching-learning process (Calder, 1990a). Calder held four public hearings across Alberta, and interviewed individuals who wished to be heard. He also interviewed individuals thought to represent various points of view. As well, he sent opinion survey questionnaires to teachers and students in 38 schools. About half the sample responded, consisting of 135 Grade 12 teachers and 238 students.

We quote the general conclusions Calder makes in the executive summary of his report.

> Although there is no consensus, and teachers see many ways in which the examinations negatively affect the teaching-learning process, they generally favor the retention of the diploma examinations. The examinations are seen more as fulfilling a political evaluative function than being of educational value. Although teachers report more negative impact of the examinations, they seem to be reluctant to dispense with them.
>
> Both teachers and students see the examinations as creating considerable stress. They question the fairness of an evaluation system where a high proportion of the student's final grade can be determined at one sitting.
>
> There is general concern that the examinations force a certain type of shallow recall learning at the cost of innovative teaching geared towards higher levels of learning.

The examinations are seen as forcing upon teachers a narrow definition of the curriculum which teachers see themselves as being forced to teach.

Both students and teachers seem rushed to complete the requirements for the examination. Insufficient time is allotted for the optional portions of the curriculum and little flexibility exists to deal with current issues. In many cases achieving high grades on the examination seems to have taken the place of mastering the subject. Although there is great pressure on students to do well on the examinations, this pressure appears to stem primarily from the pressures of university entrance. Students are seen as having to learn more and more each year to reach the same mark level of accomplishment. The basic normative nature of the examinations stresses student differentiation rather than student accomplishment.

The examinations appear to have drawn more focus on the "academic" student while ignoring the non-academic student. Because the examinations have failed to address the issue of standards for all graduating students, they have failed to address the primary concern for which they were created, that is the public's concern about the meaning of a high school diploma. By focusing on the "academic" student, the examinations have helped to support an atmosphere in the schools where the university bound students are seen as being the most worthy but approximately forty percent of high school students fail to graduate.

The examinations seemed to have helped in clarifying the instructional goals for teachers, making them more aware of the importance of assessment and how to carry out more appropriate evaluation. They have been instrumental in enforcing a standard Alberta curriculum. The examinations seem to be well developed and scored. Innovations are taking place in the nature of the examinations that will address some of their shortcomings.

The examination system is very costly to administer. Modifications could be made in the system that could still preserve its positive aspects while balancing it more with teachers' desire for a more valid assessment of the students' achievement. Consideration should be given to major changes in the system. These changes could result in major cost savings.

Teachers are concerned about the improper use of examination results and how they are used in the evaluation of their performance. Appropriate policies about the release of grades to the general public should be developed and enforced. (pp. 3-5)

Finally, we report the findings of a study commissioned by the British Columbia Ministry of Education (Anderson, Muir, Bateson, Blackmore, & Rogers, 1990, cited in Calder, 1990b). The authors found that university admissions officers did not generally use the provincial examination grades for admissions decisions, and employers did not use them in making hiring decisions. Respondents felt that final examinations yielded the beneficial result of closer adherence to provincial curriculum.

Students saw benefit in terms of increased motivation to work and study in school, but increased levels of stress was the most serious issue for them. Students also were concerned that the heavy weighting of a single test was possibly unfair. Students' test-taking skills were reported to have

increased. Students were also reported as choosing more non-examination options to reduce their workloads.

Teachers reported dissatisfaction over the narrowing of the curriculum, as well as reduced teacher morale and increased levels of stress. Teaching focussed on topics most likely to be examined, and students were reported to rely on memorization rather than critical thinking. Teachers resented the use of results to evaluate their teaching. Schools were concerned about time-tabling issues and the increased need for out-of-class tutoring for the examinations. Parents felt that schools had become more accountable. Some parents were concerned that the examination program was designed for the minority of students going on to higher education, and that a comparable examination system should be set up for the majority of students. On the other hand, parents reported a high level of confidence in the school system, unchanged as a result of implementing school-leaving examinations.

Most respondents were satisfied with the provincial examination system. Suggestions were made to reduce the weight compared to the teacher evaluations, and to improve the level of test items.

On the basis of their findings, the authors recommended that the examination program be maintained but that the Ministry should take steps to clarify issues such as the appropriate uses of the results, the relationship expected between provincial marks and school-based marks, the practice of "scaling", and the operation of the examination setting and marking committees. They also suggested that examinations should include more higher cognitive level items. Finally, they recommended that research be initiated into the questions of stress on students and teachers, the problem of examinable and non-examinable courses, the effect of provincial examinations on course content and testing practices in preceding grades, and the extent of tutorial assistance, both within and outside schools.

Both Puhlmann's (1991) study and ours constitute research at the most fundamental level called for in the 1990 Anderson et al. study cited above. The results show serious deleterious effects of the current examination system on classroom practice and environment, and they document the barriers to improving secondary school instruction. The single positive result cited in Calder's (1990) study concerns the increased adherence of teachers to the standard Alberta curriculum. Calder's findings in Alberta corroborate our findings in British Columbia.

What we find to be of particular concern is that, despite all the negative features of provincial examinations, teachers generally favour their retention. We are concerned that the longer the present system stays in effect, the more educators will become co-opted by the system and unable to envisage more productive means to achieve the goals of public education. We see this effect

in the recommendations of Anderson et al. that the present system be continued, albeit with some tinkering. We see it in the proposals for a national testing program put forward by President Bush in the United States. We do not see increased testing, as it is currently carried out, as the panacea to improve educational standards. We have spent two years observing in secondary science classrooms, interviewing teachers and administrators, and talking informally with students. We think we understand the complexity of the tasks of teaching and of student evaluation. We fear that the simplistic solutions sketched in Scenario H--let sleeping dogs lie--will become reality.

Our main concern, however, lies beyond the political ends which school-leaving examinations appear to serve. They lie with the negative impact that school-leaving examinations have on the education of young people in our society. Oakeshott has said:

> As civilized human beings, we are not the inheritors, neither of an inquiry about ourselves and the world, nor of an accumulating body of information, but of a conversation, begun in the primeval forests and extended and made more articulate in the course of centuries....Education, properly speaking, is an invitation into the skill and partnership of this conversation...(Hirst, 1975, p.53).

Our study has assessed how school-leaving examinations are affecting that conversation. In our opinion, these examinations not only fail to provide the skills necessary for young people to enter and extend that conversation, they limit and circumscribe the conversation itself.

References

Ausubel, D. (1963). The psychology of meaningful learning. New York: Grune & Stratton.

Baldridge, J.V. (1972). Organizational change: The human relations perspective versus the political systems perspective. Educational Researcher, 1(2).

Bennis, W.G., Benne, K.D., Chin, R., & Corey, K.E. (1976). The planning of change (3rd ed.). New York: Holt, Rinehart and Winston.

British Columbia Ministry of Education (1983). Ministry Policy Circular. Victoria, B.C.: Author

Bruner, J. (1961). The Process of Education. Cambridge, Mass: Harvard University Press.

Calder, P. (1990a). Impact of diploma examinations on the teaching-learning process. Edmonton: Alberta Teachers' Association.

Calder, P. (1990b). Impact of external examinations on the teaching-learning process: A review of the literature. Edmonton: Alberta Teachers' Association.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). Graphical methods for data analysis. Belmont, CA: Wadsworth.

Darling-Hammond, L, & Wise, A. E. (1985). Beyond standardization: State standards and school improvement. The Elementary School Journal, 85(8), 315-336.

Goodlad, J.I. (1984). A place called school: Prospects for the future. N.Y.: McGraw Hill.

Gronlund, N. E. (1981). Measurement and evaluating in teaching (4th ed.). New York: Macmillan.

Hirst, P. (1975). The logic of the curriculum. In M. Golby, J. Greenwold, & R. West, Curricular design. London: Croom Helm.

Hoffman, B. (1962). The tyranny of testing. New York: Collier.

Joyce, B., & Weil, M. (1980). Models of teaching. Englewood Cliffs: Prentice Hall.

MacKinnon, A.M. (1989). Conceptualizing a reflective practicum in constructivist science teaching. Unpublished doctoral thesis, University of British Columbia, Vancouver.

Morgenstern, C. F., & Renner, J. W. (1984). Measuring thinking with standardized science tests. Journal of Research in Science Teaching, 21(6), 639-648.

Naylor, C.S. (1990). Dropping out of high school: An exploratory and critical analysis. Unpublished master's thesis. Simon Fraser University, Burnaby, B.C.

Popham, W. J. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 68(9), 679-682.

Popham, W. J., Cruse, K. L., Rankin, S. C., Sandifer, P. D., & Williams, P. L. (1985). Measurement-driven instruction: It's on the road. Phi Delta Kappan, 66(9), 628-634.

Professional Standards for Teaching Mathematics. (1991) . Reston, VA: National Council of Teachers of Mathematics.

Puhlmann, K. (1991). Interpreting the mathematics curriculum. Unpublished doctoral dissertation. University of Alberta, Edmonton, AB.

Sewall, G.T. (1983). Necessary lessons: Decline and renewal in American schools. N.Y.: The Free Press.

Shepard, L. A. (1989). Why we need better assessments. Educational Leadership, 46(7), 4-9.

Shulman, L., & Tamir. (1973). Research on Teaching in the Natural Sciences. In R. M. W. Travers (ed). Second Handbook of Research on Teaching.Chicago: Rand McNally. 1098-1148

Stodolsky, S.S. (1988). The subject matters: Classroom activity in mathematics and social studies. Chicago: University of Chicago Press.

Sullivan, B. M. (1988). A legacy for learners: The report of the royal commission on education. Victoria, BC: Province of British Columbia.

Van Der Gegt, R.& Knip, H. (1990) Implementing mandated change: The schoool as change contractor. Curriculum Inquiry. 20(2) 183-203.

Van Til, W., Ed. (1976). Issues in secondary education. (NSSE Yearbook: Part 2). Chicago: University of Chicago Press.

Wideen, M., Pye, I., Naylor, C. & Crofton, F. (1990). A platform for change: A study of Surrey secondary schools. Burnaby: Simon Fraser University Press.